Table of Contents

Contract No. 53-3198-8-41
Full and open competition

Contract Amount $178,959

Treatment of Incomplete and Out-of-Scope Case Reviews in Food Stamp Quality Control

Prepared for:

Office of Analysis and Evaluation
U.S. Department of Agriculture
Food and Nutrition Service
Alexandria, VA 22302

June 12, 1989

WESTAT
An Employee-Owned Research Corporation

Table of Contents

WESTAT
An Employee-Owned Research Corporation
1650 Research Blvd • Rockville, MD 20850-3129 • 301 251-1500

June 12, 1989

Steven Carlson
Office of Analysis and Evaluation
U.S. Department of Agriculture
Food and Nutrition Service
3101 Park Center Drive, Rm. 215
Park Center Office Building
Alexandria, VA 22302

Dear Steve:

Transmitted herewith are 20 copies of the revised version of what we submitted earlier as the final report on the Treatment of Incomplete and Out-of-Scope Case Reviews in Food Stamp Quality Control, prepared under contract number 53-3198-8-41.

This revised version of the report was prepared in response to your request for a more explicit summary of the treatment of the uncompleted cases in the estimation of error rates. The more explicit detail appears at the end of Section 2, and also in a revision of the Summary and Recommendations section. Also, we have taken this opportunity to do some minor editing. In accordance with the contract we are sending a separate copy to Elaine Lynn, the Contracting Officer.

We appreciated receiving and have tried to be responsive to the comments from you and your staff on the first draft, and the additional request that led to this revised final version.

Sincerely,

Morris

Morris H. Hansen
Chairman of the Board

MH/jsn
Enclosures

Table of Contents

Contract #53-3198-8-41 $178,959 Awarded under open bid

Treatment of Incomplete and Out-of-Scope Case Reviews in Food Stamp Quality Control

by

Morris H. Hansen and Benjamin J. Tepping

June 12, 1989

Submitted to

Mr. Steven Carlson
Contracting Officer's Representative
Office of Analysis and Evaluation
Food and Nutrition Service
Department of Agriculture
Room 212
3101 Park Center Drive
Alexandria, VA 22302

Westat, Inc.
1650 Research Boulevard
Rockville, Maryland 20850

Treatment of Incomplete and Out-of-Scope Case Reviews
in Food Stamp Quality Control

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

This paper is concerned with the treatment of active cases in the Food Stamp Quality Control samples that do not now receive review, either by the state or by the regional FNS office, even though the case received food stamps in or for the sampled month. Some of these cases are not subject to review under current rules. Specifically, these cases include cases in which all members of the household have died or moved out of the state, cases under investigation for fraud, and cases pending a hearing. Others for which a review is intended but not completed are cases in which the recipient could not be contacted for a personal interview or refused an interview.

Any error made in the determination of eligibility or coupon allotment for such cases is therefore not directly reflected in the estimates of payment error rates based on the samples. They are implicitly treated as having the same average error characteristics as the average for all completed reviews.

Under current practice the proportion of sample cases not reviewed averages nearly ten percent over the United States, and has been as high as twenty-five percent in some states. These not-reviewed cases include an unknown number of cases not active in the sample month. The procedures described in this report would identify and greatly reduce the frequency of the active cases for which reviews are not completed.

We recommend that the only cases selected for the sample that are not subject to review be cases that were not active in the sample month and cases that are dropped from the sample at random to avoid a sample size greater than that required.

We recommend that the disposition code be revised to distinguish four disposition classes of cases selected for the state sample, namely

Table of Contents

(1) Cases reviewed by the state;

(2) Cases not active in the sample month;

(3) Cases dropped at random to reduce the size of the state sample;

(4) Other active cases not reviewed.

The Federal subsample is to be a subsample of eligible cases in the state sample; that is, it is to be a subsample of cases of the classes (1) and (4). In addition, the Federal reviewer should check the disposition of all cases in the state sample whose review was not completed. Note that the only cases that are properly dropped from the state sample are cases in classes (2) and (3). If the Federal review of the state disposition finds cases improperly classified as in class (2) or class (3), such cases should be added to the list of eligible sample cases and the subsampling extended to provide their contribution to the Federal subsample. We show in detail how the results should be incorporated in the estimates of the payment error rate for the state.

We anticipate that adoption of the recommended procedures would mean that many or most of the reviews of active cases currently recorded as not subject to review or not completed would be completed by reference to the case records and to other sources of information. However, some uncompleted reviews of active cases may remain. We describe several alternatives for the treatment of such cases. The choice of imputation procedures for such cases requires additional research and therefore cannot be implemented immediately. Moreover, if the proportion of cases in class (4) is very small, it may not be worthwhile to impute for incomplete reviews.

We recommend that the detailed characteristics of sample cases, whether their review is completed or not completed, be retained in the microdata file. Such data will be useful for future studies of the bias of non-response and of alternative imputation procedures, and for imputation for nonresponse.

We recommend that an investigation be undertaken of a sample of remaining not-completed state reviews to study the incidence of eligibility

Table of Contents

and allotment errors in such cases, and whether explicit imputation is worthwhile for these remaining cases, and, if so, how imputation should be done.

1. INTRODUCTION

The Food and Nutrition Service (FNS), Department of Agriculture, is responsible for national administration and management of the Food Stamp Program. Each state is responsible for administration within the state and for day-to-day operation of the program, including accepting applications for benefits, investigating and determining eligibility of applicants, determining the amount of allotments to applicants found to be eligible, and issuing benefits. Only the costs of administration within a state are shared by the state and the Federal government, ordinarily each paying approximately half.

The quality control (QC) system is prescribed by law and by agency regulations and is in operation in each of the states. Each state selects and reviews a monthly sample of active cases that are identified, under current QC regulations, as in scope for QC review. With certain exceptions, active cases are households issued benefits for the sample month. The out-of-scope cases are described below. The state review is an intensive re-investigation of each in-scope active sample case, including a required personal interview, to determine the eligibility and the amount of the benefit or allotment for the case. Efforts are made by the FNS to have the states complete the reviews of substantially all the in-scope cases selected for the state sample. The observed completion rate for in-scope cases is usually over 95 percent, and is quite often 98 percent or more.

The total payment error rate for a state is the average payment error, including both overpayment and the absolute value of underpayment errors, divided by the estimated average payment. The overpayment error rate is equal to the estimated average overpayment (underpayments being treated as zero overpayments) divided by the estimated average payment. The overpayments include both payments to eligible cases that are above what they should be, and payments to ineligible cases. An underpayment error rate is also computed for cases receiving benefits, defined as the ratio of the average underpayment (overpayments being treated as zero underpayments) to the average payment. Case error rates are also estimated. These include the proportion of active cases that are ineligible, the proportion with overpayments, and the proportion with underpayments. Also, samples

Table of Contents

of terminations and denials are reviewed, and the proportion of cases with improper terminations and the proportion of applications that are improperly denied are estimated. The estimated underpayment error rate does not include any cases that were improperly denied benefits.

This paper is concerned with the treatment of active cases in the Food Stamp Quality Control samples that do not now receive review, whether by the state or by the regional FNS office, even though the case received food stamps in or for the sampled month. Specifically, these cases include cases in which all members of the household have died or moved out of the state, cases under investigation for fraud, cases pending a hearing, and cases in which the recipient could not be contacted for a personal interview or refused an interview. Any errors made in the determination of eligibility or coupon allotment for such cases are therefore not directly reflected in the estimates of payment error rates based on the samples. They are implicitly treated as having the same average error characteristics as the average for all completed reviews.

The General Accounting Office in a limited investigation[1] of active cases not receiving a review concluded that they differ substantially from those receiving a complete review, and tend to have considerably higher error rates than the cases for which reviews are completed. In this study we examine the treatment of such cases and suggest some solutions.

In Section 1 we describe more fully the background in which the problem arises. In Section 2 we present a possible treatment that involves explicit inclusion in the estimates of cases not now reviewed. In Section 3 we discuss imputation techniques that may be employed when cases cannot reasonably be completed by the procedures suggested in Section 2.

---

[1] General Accounting Office, *Food Stamp Program: Refinements Needed to Improve Accuracy of Quality Control Error Rates*, Report to the Chairman, Committee on Governmental Affairs, United States Senate. September 1986.

1.1 Structure of the state sample for active cases

The Food Stamp Quality Control Program calls for each state to select samples of its active cases and to review each in-scope sample case, to determine whether the action was in accord with established policy and guidelines. The design of the sample is proposed by the state, but must be approved in advance by the Food and Nutrition Service (FNS), consistent with the FNS regulations that specify the minimum size of the sample for that state as well as other requirements for the sample design. The states have an option to reduce (or increase) the sample size under specific conditions set by Federal regulation but only to specific levels. The state selects and reviews the QC sample, and submits the selected sample, along with the results of each case review, to the regional office of FNS. Present practice is to require that this review include a personal interview with the recipient. The regional office selects a subsample of the completed reviews in the state sample, and conducts a re-review of each case in the subsample. This review is dependent, in that it makes use of the information already obtained in the state review, but may also involve the collection and use of additional information. The Federal re-review does not require a personal interview. Cases for which the Federal determination differs from the state determination are reviewed with the state before the final Federal determination is fixed. As discussed later, the regional office reviews each case in the state QC sample that did not receive a complete review by the state, and may refer some of these cases back to the state for further action. An arbitration procedure resolves differences of opinion with respect to the Federal determinations.

The QC program has multiple objectives. The annual state sample, whose minimum size varies by state from about 250 to about 3000, not only provides estimates of the proportion of incorrect decisions and the percentage of benefit dollars under-paid and over-paid to active cases, and the total payment error rate, but also yields, to the state, useful information on the sources of errors. Such information can lead, and has in the past led, to corrective action including changes in modes of administration and to policy changes.

Table of Contents

The Federal subsample, whose size varies by state from about 170 to about 400 completed reviews, is used to monitor the quality of the state administration of the QC program and to provide estimates of the state under- and overpayment error rates and of the total payment error rate. The monitoring function is guided principally by the degree to which state and Federal determinations for individual cases differ, and by the character of those differences. The computed error rates are based on the Federal subsample reviews but, through the use of the double-sampling regression estimator, also make use of the state reviews of the larger state sample to increase precision substantially. The regression estimator is fully described and evaluated in an earlier report.[2] The resulting error rates permit state-by-state comparison of error levels and thus provide guidance on which states should be given special attention. We note that the sizes of the state samples and the Federal subsamples actually achieved in completed reviews differ somewhat from the planned sizes. We are concerned here especially with the cases receiving benefits for which reviews were not completed, since these relate to possible biases.

The estimated overpayment error rate, after a minor adjustment by a measure of the extent to which the state failed to complete reviews on all sampled cases, has also been a tool for holding states accountable for the accuracy of their determinations of household eligibility for food stamps and the food stamp benefits that are issued. States whose adjusted overpayment error rates exceed a specified threshold have been assessed a fiscal liability intended to recover a portion of the overpayment. States with low estimated error rates, including error rates for underpayments and negative actions, have been eligible for additional Federal funding. The recently enacted Hunger Prevention Act of 1988 bases the state liability on the total error rate rather than on the overpayment error rate.

The references to "current" procedures in this report refer to the procedures of the QC system that were in effect during the 1987 fiscal year. The 1988 Act, in addition to making the state's fiscal liability depend on the sum of the underpayment and overpayment error rates, establishes

---

[2]Hansen, M.H., and Tepping, B.J. (1987), A Statistical Evaluation of Food Stamp Quality Control, Westat, Inc., Rockville, Maryland.

Table of Contents

thresholds for the total payment error rate. The procedures proposed in this report are applicable both to the procedures applied in 1987 and to the provisions of the 1988 Act.

1.2 Design of the state sample

For each state, the minimum size of the annual state sample is specified by FNS. The state selects a portion of the sample each month, so that the annual sample is stratified by month. Some states also stratify within each month, for example by region of the state, so as to provide separate estimates for two or more regions. Some states define as a stratum those cases that also receive benefits from Aid to Families with Dependent Children (AFDC), and draw the Food Stamp QC sample to overlap the AFDC-QC sample. Then, a single review serves both QC programs as a means of reducing administrative costs.

The QC sampling rate in a state may vary from stratum to stratum. The sampling rate used for a given month is based on estimates of the annual caseload, and thus may be revised later in the fiscal year to avoid sample sizes below the required minimum or substantially above that minimum. In estimating variances, the sample of active cases within each stratum is treated as a simple random sample, although it is typically a systematic sample. If the organization of the case record files from which the systematic sample is drawn is such that the serial correlations in the files are low, as we presume they are, systematic sampling is substantially equivalent to simple random sampling. Also, the stratification by months is ordinarily not reflected in the variance estimator unless differential sampling fractions have been used. This deviation from exact application of theory is not deemed to be important, as is demonstrated in the Westat report referred to above.

1.3 Design of the Federal sample

The Federal subsample is a systematic subsample of the State sample, drawn in the order in which reviewed cases are received from the

Table of Contents

state. The Federal subsampling fraction varies from about a third or more of the state sample in states with relatively small state samples, to about a sixth in states with relatively large state samples. Thus, within a state the probability of selection is proportional to the probability with which the case was selected for the state sample. The stratification by month is usually not explicitly reflected by the estimator, but other stratification that is employed is reflected when the state has identified and used varying sampling fractions by strata in the state sample. The full state sample results are used in the regression estimator along with the results from the Federal review. In this use no changes are permitted in the state sample results for any completed state reviews that have had a chance of being selected for the Federal sample. This is a necessary restriction to keep the regression estimator unbiased. Of course, appropriate changes can be made for other analytic purposes.

1.4 Scope of this report

This report is concerned with the treatment of incomplete sample cases and sample cases not subject to review under the current rules. The state selects a sample from its file of presumed active cases each month. However, some sampled cases are in fact not active, i.e., did not receive food stamps in or for the month in question. These are not subject to review and no review is attempted. In addition, there are active cases that are classified by the state as not subject to review for which, after certain initial information is obtained, no review is attempted. These include cases in which all the members of the household have either died or moved out of the state, cases which are under investigation for fraud, and cases which receive continued benefits pending a fair hearing. The state attempts to review each sample case that is subject to review but may not succeed in completing the review. This may occur because, for example, a personal reinterview cannot be completed because the household cannot be located, or the case record cannot be located, or the household members refuse to be interviewed. The purpose of this report is to suggest possible treatments for the active sampled cases now classified as not subject to review, and for uncompleted reviews of cases subject to review. The current practice of requiring a personal interview for each completed sample case in the state review contributes to the number of

a nonresponse imputation procedure that takes advantage of whatever limited information is known for the case. Some procedures for imputation of state findings for any remaining uncompleted cases are considered in Section 3 of this report.

Table of Contents

Fig. 1. Number of states by percent of cases with disposition code 1, for fiscal years 1986 and 1987

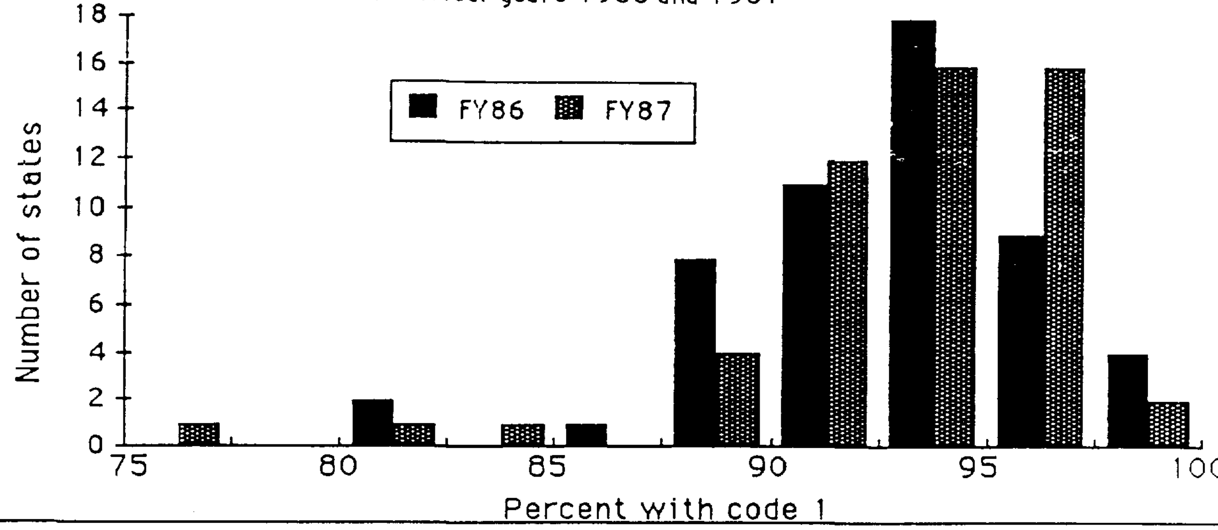

Fig. 2. Number of states by percent of cases with disposition code 2, for fiscal years 1986 and 1987

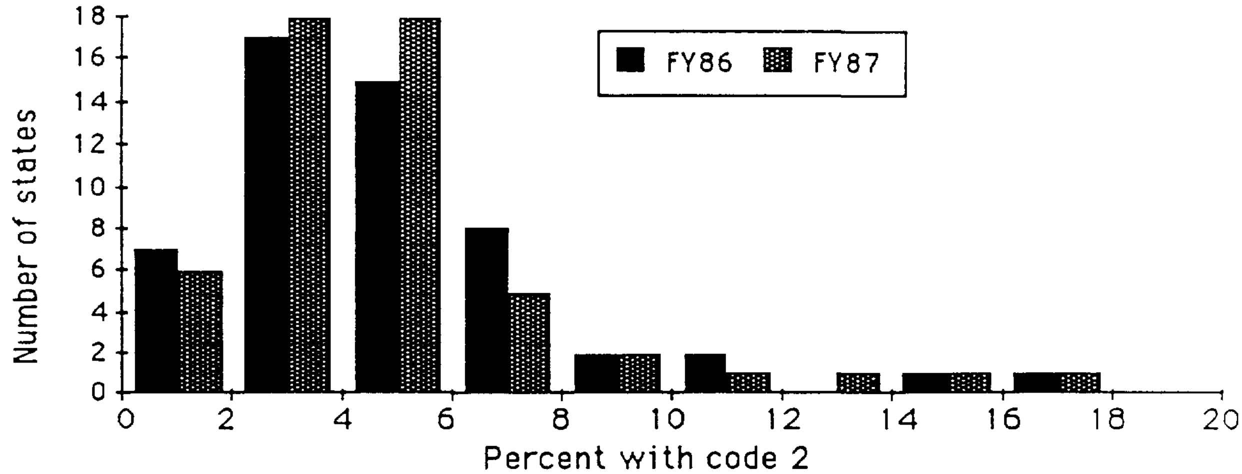

Fig. 3. Number of states by percent of cases with disposition codes 4-7, for fiscal years 1986 and 1987

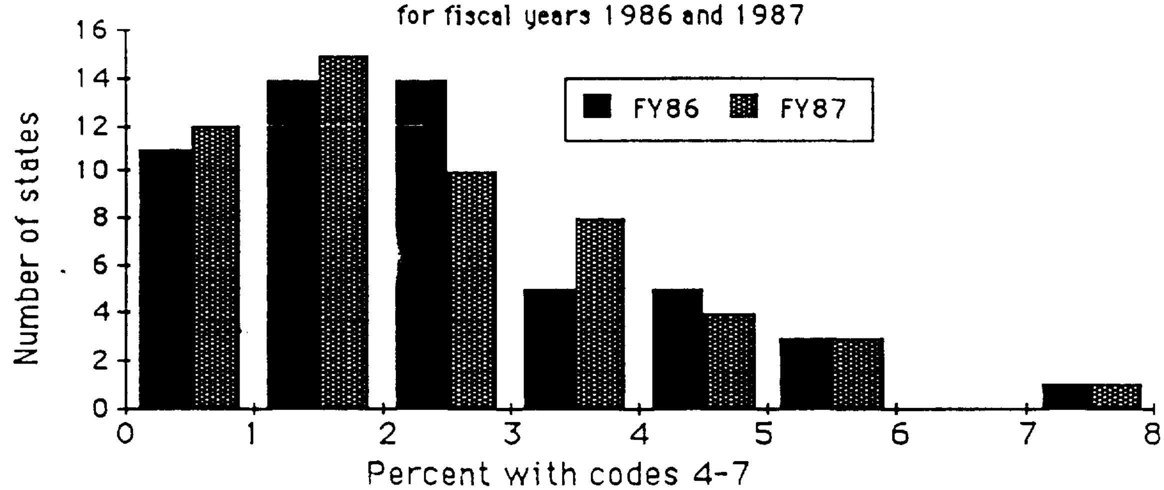

Table of Contents

2. PROPOSED TREATMENT OF CASES NOT NOW SUBJECT TO REVIEW OR NOT COMPLETED

In the current Food Stamp practice, cases selected for the state sample but for which reviews are not completed by the state QC are of three classes:

(a) Cases not active in the sample month (code 2);

(b) Active cases classified as not subject to review (also code 2); and

(c) Other active cases considered as subject to review whose review was not completed (codes 4 to 7).

Data for the class (a) cases play no role in the regression estimator of payment error rate, nor should they. The estimator has the form $\bar{y}'/\bar{u}$, where

$$\bar{y}' = \bar{y} + b\,(\bar{X} - \bar{x}) \tag{1}$$

and $\bar{X}$ denotes the mean payment error determined by the state review for all completed cases in the full state sample, $\bar{u}$ denotes the mean payment for all completed cases in the full state sample, and $\bar{x}$ and $\bar{y}$ denote respectively the state and Federal mean payment error for cases in the Federal subsample for which <u>both</u> state and Federal reviews were completed.

As indicated above, in current practice, all cases in the Federal subsample whose state review was completed are reviewed by the Federal reviewer. The Federal reviewer may reclassify some of these cases as not subject to review and therefore in classes (a) or (b). In addition, the Federal reviewer examines the disposition assigned by the state reviewer for every case in the state sample in classes (a), (b), and (c) above. Instances in which the Federal review of such cases disagrees with the state disposition are referred back to the state, which may then complete a review and change the disposition. Any final disagreements are resolved by arbitration. The Federal reviewer may also not succeed in completing a review of some cases in the Federal subsample of cases completed by the state, analogous to class (c). Also, for some cases in classes (b) or (c) the review is completed by the Federal reviewer, whose finding becomes the state finding; the case is treated as completed by the state.

The only data currently permitted to enter into the computation of $\bar{y}$ and $\bar{x}$ are for the cases whose review is completed by both the state and the Federal reviewer. The existence of active sample cases that are not permitted to affect the estimate of the payment error rate results in biased estimates, in general. This fact gives rise to several important questions, discussed below.

How should the current class (b) be handled in the state review?

The state sample cases currently classified by FNS as in class (b), that is, active cases not subject to review, are of four distinct types. One type consists of randomly subsampled cases that were dropped from the original sample selection to avoid a sample larger than that required. Such cases are not reviewed and, of course, errors of eligibility or allotment in such cases do not and should not enter into the computation of $\bar{y}'$. A second type consists of cases that are not "regular," that is, cases that received payments under disaster certification, cases receiving restored benefits, and cases that were participants in demonstration projects. Such cases should be reviewed only to the extent that they are also receiving benefits as regular cases. For example, restored benefits should not be considered but only eligibility and allotment under regular rules for the sample month. A third type consists of cases in which a review may interfere with other activities; these include cases that are under investigation for violation of rules or for fraud, and cases which are receiving continued benefits pending a fair hearing. The review for these cases could be handled by omitting, if policy requires it, those parts of the standard review process that would interfere with the other procedures. While the findings may not be regarded as definitive as those given by the standard review, this partial review would presumably result in a smaller contribution to the bias of the error rate than would their exclusion completely. A fourth type consists of cases for which personal interviews have not been completed because all members of the household have died or moved out of the state. Error determinations for these cases should be made by doing as complete a review as is possible on the basis of the case record and collateral sources of information.

Under the procedure we propose, as in current practice, the Federal reviewer will examine the disposition of every case in the state sample whose review was not completed by the state and may request the state to correct the disposition and to complete the case. Under the procedures proposed here, however, the number of such non-completed reviews will be greatly reduced. Changes in the $\bar{X}$, $\bar{y}$, and $\bar{x}$ may result from such 100 percent review. The validity of the regression estimator depends on the fact that $\bar{x}$ is an unbiased estimate of $\bar{X}$, which will follow from this procedure. Thus, the state disposition and finding for a case must be the same whether or not a case in the state sample is selected for the Federal subsample.

How should the current class (c) be handled in the state review?

This class (c) consists of cases in which the household could not be located for a personal interview or the household refused to give a personal interview. Thus, the standard review could not be completed. The current procedure is to exclude them from the computation of $\bar{X}$ and $\bar{x}$. This procedure implicitly imputes findings for them as the average of the reviews that are completed. Since some information is available for such cases in the case records and other information can sometimes be obtained by telephone interviews, we suggest that a more satisfactory alternative is to make the finding on the basis of the best information available or attainable and to omit a case only if a reasonable finding cannot be made. The remaining uncompleted reviews can be implicitly imputed at the average for the sample of completed cases if they are simply omitted from the computations. However, better methods of imputation may be used, especially if the number of these is still moderately large. To gauge the usefulness of alternative imputation methods it would be desirable to have a great deal more data on the characteristics of these cases than is now readily available.

As in the present class (b), the Federal reviewer may suggest a change in the state disposition of a case, which may then become a completed review under the suggested procedure. Again, we believe that the number of such cases will be greatly reduced.

Table of Contents

How should the current class (b) be handled in the Federal review?

This question divides into two: how should a case be treated in the computation of $\bar{y}$ and how should it be treated in the computation of the regression coefficient b?

We consider first the computation of $\bar{y}$. We note that the regression estimator remains valid even if the cases that enter into $\bar{y}$ are not exactly the cases that enter into $\bar{x}$. The only effect of different cases being used for those two statistics is to decrease the correlation between $\bar{y}$ and $\bar{x}$ and thereby to increase the variance of $\bar{y}'$. The increase in variance ordinarily will be small if nearly all the cases contributing to $\bar{y}$ also contribute to $\bar{x}$. Since the value that is to be estimated is the error rate that would result from a complete Federal review of the active caseload (preceded by a state review), we suggest that the determination of eligibility and allotments for cases in class (b), for purpose of computing $\bar{y}$, be made by the Federal reviewer using the procedures and criteria we have given above for the state reviewer.

In some instances, the Federal reviewer may request the state to augment the state review for a case. The additional review by the state may change the disposition of the case. We note that, as the result of the state changing the state disposition of a case to "completed," the case enters into the computation of $\bar{X}$. If that case is also selected for the Federal subsample, it also enters into the computation of $\bar{x}$ and $\bar{y}$.

The problem is a little different for the computation of b. That computation involves the sum, over the Federal subsample, of products of the state and Federal error findings. Thus, if a case has finally been classified as in class (b) *either* by the state review or the Federal review, the case should not enter into the computation of the regression coefficient b. We suggest also that extreme (i.e., very low) values of b be avoided by setting a minimum acceptable value of b on the basis of state-specific data for several recent periods and empirical knowledge of the sampling distribution of estimates of the correlation coefficient. This is because low values of the estimate b are most likely the result of sampling variance unless the true correlation between state and Federal error findings is very low, which we believe is

Table of Contents

unlikely. However, we suggest that until sufficient data are available and analyzed, no minimum criterion should be applied.

How should the current class (c) be handled in the Federal review?

The same considerations that apply for class (b) apply here. To the extent that imputations are used for the Federal error finding, the imputations can make use of data acquired by the state review as well as the state imputations (if any).

Concluding comments

The consequence of the above recommendations, if adopted, would mean that in the future, class (b) as now defined would disappear. The other cases now in class (b) would become completed reviews or would be transferred to (a) or (c). Also most of the cases currently in class (c) would become completed reviews. It would be proposed to impute for any remaining class (c) cases, but procedures for doing so remain to be examined and developed.

The consequence of the state QC review would then be to assign one of the following four disposition codes to the cases selected for the state QC sample.

(1) Cases reviewed by the state;

(2) Cases not active in the sample month;

(3) Cases dropped at random to reduce the size of the state sample;

(4) Other active cases not reviewed.

If desired, subclasses could be assigned within class 4 to show reasons the review was not completed.

Table of Contents

The Federal subsample is to be a subsample of eligible cases in the state sample; that is, it is to be a subsample of cases of classes (1) and (4). In addition, the Federal reviewer should check the disposition of all cases in the state sample whose review was not completed. Note that the only cases that are properly dropped from the state sample are cases in classes (2) and (3). If the Federal review of the state disposition finds cases improperly classified as in class (2) or class (3), such cases should be added to the list of eligible sample cases and the subsampling extended to provide their contribution to the Federal subsample.

As in current practice, the Federal reviewer may request the state to review cases in class (4), or may complete the review or imputation of such cases. Nevertheless, there may remain unreviewed cases of that class. Imputation of eligibility and payment may or may not be implemented for these. If implemented, the values of $\bar{X}$, $\bar{x}$, $\bar{y}$ and b in the regression estimator are to be computed as in current practice. That is, $\bar{X}$ is the mean error in the state sample as determined by the state review or imputed, $\bar{x}$ is the mean error in the Federal subsample as determined by the state review or imputed, $\bar{y}$ is the mean error in the Federal subsample as determined by the Federal review or imputed, and b is the regression coefficient computed from the Federal subsample values determined by the state and Federal reviews (again including imputations).

However, if all reviews are not completed and if imputation is not fully implemented there will remain uncompleted reviews. In this event, the proposed estimation procedure is as follows:

n = the number of active cases in the state sample with completed state reviews, including any imputations—i.e., the number of cases assigned disposition code (1);

n' = the number of cases from among the n that are selected for the Federal subsample;

$\tilde{n}'$ = the number of cases in the Federal sample for which both the Federal and state reviews were completed (including any imputations);

Table of Contents

$\hat{n}'$ = the number of cases in the Federal subsample for which the Federal review was completed (including any imputations and possibly including some cases for which a state review was not completed).

The mean error per case is estimated by

$$\bar{y}' = \bar{y} + b\left(\bar{X} - \bar{x}\right).$$

The statistics in this formula are computed as follows.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\bar{x} = \frac{1}{n'}\sum_{i=1}^{n'} x_i$$

$$\bar{y} = \frac{1}{\hat{n}'}\sum_{i=1}^{\hat{n}'} y_i$$

$$b = \frac{\sum_{i=1}^{\tilde{n}'} (x_i - \tilde{x})(y_i - \tilde{y})}{\sum_{i=1}^{\tilde{n}'} (x_i - \tilde{x})^2} = \frac{\sum_{i=1}^{\tilde{n}'} x_i y_i - \tilde{n}'\,\tilde{x}\,\tilde{y}}{\sum_{i=1}^{\tilde{n}'} x_i^2 - \tilde{n}'\tilde{x}^2}$$

where

$$\tilde{x} = \frac{1}{\tilde{n}'}\sum_{i=1}^{\tilde{n}'} x_i$$

$$\tilde{y} = \frac{1}{\tilde{n}'}\sum_{i=1}^{\tilde{n}'} y_i\,.$$

Table of Contents

3. IMPUTATION FOR SAMPLE CASES NOT REVIEWED

We note that the only unbiased procedure for dealing with active cases in the sample which are not reviewed is to complete the necessary work for all of them. This is unbiased if accurate information is collected and appropriate action taken. Of course, this is a theoretical goal that cannot be completely achieved in practice, and it is necessary to do at least a limited amount of imputation. However, if the proposals given above are adopted the frequency of uncompleted case reviews will be reduced greatly.

Sometimes, the imputation is implicit rather than explicit, for example by treating the completed interviews as the full sample. That is the procedure now used. This procedure implicitly imputes to non-reviewed cases the average error determinations of those for which completed results were obtained. There is much evidence in various studies that this particular assumption is likely to be in error and difficult to defend, although if the completion rate is high enough, as is expected under our recommended changes in procedures, the results may be reasonably acceptable. It is likely that such implicit imputation can be improved by an imputation procedure that takes account of the information available for both reviewed and non-reviewed cases. In Food Stamp QC there is usually a considerable amount of such information.

We first elaborate briefly some alternative imputation procedures to be considered in obtaining state estimates adjusted for missing data. We then suggest a possible procedure for imputing for any missing data in the Federal subsample, given that complete information is available for the active state sample cases, either directly or by imputation. These state and Federal results can then be applied in a regression estimator as now, but with information included in each sample for all active sampled cases. The procedures described below are preliminary and illustrative, and likely may be modified in various ways as work proceeds and specific results are obtained, and analyses completed and evaluated.[3] These procedures cannot

[3]For descriptions and evaluation of the issues and procedures for dealing with incomplete data, see *Incomplete Data in Sample Surveys*, Vols. 1, 2, and 3, Panel on Incomplete Data of the Committee on National Statistics, National Research Council, Academic Press (1983). Also see:

Kalton, G. (1983), *Compensating for Missing Data*, Ann Arbor, MI: Institute for Social Research; and

Table of Contents

be specified completely in this report because the new recommended procedures will greatly reduce the number (and possibly the characteristics) of cases with missing data.

(1) Average Imputation by Error-Prone Subclasses. In this approach the completed cases in the state sample are subclassified into a number of more or less homogeneous subclasses that can be defined by the base information, that is, the information that is available in the records for both the completed cases and the remaining uncompleted cases. The base information can include any of the information provided by the case records for the state sampled cases. The most common approach is to identify a few of the base information characteristics that are especially related to the existence or the magnitude of payment error. These will be characteristics that are correlated with the observed payment errors. These may include characteristics such as education, income or lack of income of household members, number of persons in household by age class, marital status and pregnancy, employment status of household members, place of residence, liquid assets such as bank accounts, AFDC eligibility, physical handicaps, the magnitude of the issuance, etc.

The goal is to achieve as much homogeneity within the subclasses or cells as feasible with respect to eligibility, appropriate issuance, and error proneness. Our preliminary investigations of techniques for defining strata by means of the SEARCH algorithm, discriminant analysis, and multiple regression have suggested the possibility of using such strata as error-prone groups for use in imputation. Their effectiveness can be evaluated after experience shows the percentages to which uncompleted reviews descend after implementation of the procedures suggested above.

Given such nonresponse cells, the adjustment for nonresponse in the state findings might be accomplished as described below. (This analysis is for overpayment errors, but is extendible to underpayment errors.)

Little, R.J.A. (1982), *Models for Nonresponse in Sample Surveys*, Journal of the American Statistical Association, 77, 237-250.

Rubin, D.B. (1987), *Multiple Imputation for Nonresponse in Surveys*, John Wiley & Sons, New York.

Table of Contents

Let

$$\bar{X}_c = \frac{\sum^{n_c} W_{ci} X_{ci}}{\sum^{n_c} W_{ci}}$$

be the average overpayment error in cell c for all completed cases (including the additional completes obtained by reference to the case records), where, for this purpose underpayment errors are treated as zero overpayments. The n_c is the number of completed cases in cell c, and the W_{ci} are the reciprocals of the probabilities of selection in the original sample. The W_{ci} will all be equal unless variable sampling fractions have been used in selecting the initial state sample.

The uncompleted cases are classified into the same cells as the completed cases, and the value $\bar{X}_c$ is the imputed overpayment error for each uncompleted case in the cell. After these imputations are made the adjusted average of the state error findings (including the effect of imputations) would be obtained as

$$\bar{X}_a = \frac{\sum_c^M \sum_i^{\tilde{n}_c} W_{ci} X_{ci}}{\sum_c^M \sum_i^{\tilde{n}_c} W_{ci}}$$

where $\tilde{n}_c$ is the number of active cases in cell c, whether originally complete or incomplete, and M is the number of cells.

The Federal subsample cases are then identified.

The average of the state findings, after imputation, for cases in the Federal subsample, is then

$$\bar{x}_a = \frac{\sum_{c}^{M} \sum_{i}^{\tilde{n}'_c} W_{ci} X_{ci}}{\sum_{c}^{M} \sum_{i}^{\tilde{n}'} W_{ci}}$$

where $\tilde{n}'_c$ is the number of active cases in cell c in the Federal subsample.

The next step is to impute the Federal finding to each of the cases in the Federal subsample for which a Federal finding is missing. Similar procedures to those presented above for state imputation may be used for the imputation of missing values in the Federal review. However, since relatively high correlations have generally been observed between state findings and Federal findings, we tentatively suggest a very simple regression procedure for the Federal imputation.

Compute the simple regression of the Federal findings, Y_i, on the state findings, X_i,

$$y = a + bx$$

where the a and b are estimated by standard linear regression, and using only the cases with completed Federal findings (including any cases that have imputed state findings). This equation is then used to provide an imputed Federal finding for the individual cases in the Federal subsample with a missing Federal finding. After this imputation is completed, we obtain the average of the Federal findings

$$\bar{y}_a = \frac{\sum_{i}^{M} \sum^{\tilde{n}'_c} W_{ci} Y_{ci}}{\sum^{M} \sum^{\tilde{n}'_c} W_{ci}}$$

where the Y_{ci} include any imputed findings.

The usual regression estimator is then applied to the Federal findings, that is

$$\bar{y}'_a = \bar{y}_a + b(\bar{X}_a - \bar{x}_a) .$$

It will be desirable to explore alternatives for estimation of the regression coefficient in the presence of missing data. Some sensitivity analyses should be used to evaluate the effectiveness of this procedure, and of some possible variants of it. The procedures should be extended to stratified sampling.

(2) Imputation of Missing State Findings Using a Hot Deck Procedure. The hot deck procedure could be applied in various ways, and may have some advantages and some disadvantages that should be examined after the new procedures have been implemented and data become available. We describe a hot deck procedure that begins with the cells for the state sample obtained as identified for imputation procedure (1) described above.

However, instead of imputing the average for a cell, a "best" matching case within the cell would be identified for each case with missing state data. This could be accomplished by sorting both the cases with completed and with missing data into cells on the basis of one or more characteristics (e.g., on the amount of issuance and size of household), and using within a cell a neighbor next in the sort to each case with missing data as the donor. Then impute for that case the same information as observed for the donor, except that no donor would be used for more than one imputation.

Rather than by sorting, the "nearest neighbor" could be defined by a distance function in terms of several characteristics of the cases, like the procedure used in factor analysis. The rest of the procedure, including imputation for the missing Federal findings, will be as described above.

This hot deck procedure has the advantage of possibly reducing the bias due to imputations over the average cell imputation procedure, described in (1) above, but may increase the variance from imputation. The increase in variance would be relatively minor with the low levels of imputation that we anticipate. If needed, any increase in variance could be substantially reduced by using multiple imputations, as suggested by Rubin (cited above). This could be explored and evaluated when data become available. To state it briefly, the method uses the data from more than one donor case to reduce bias in the estimated variance.

(3) Regression Imputation. Another procedure that could be considered for imputation of missing state findings is the application of a multiple regression model. Separate regressions might be computed for each of several subgroups of the state sample that would be identified experimentally, with the hope of improving the model fit, and taking some account of some interactions.

In this procedure the independent variables used in the multiple regression equation would be identified in a stepwise procedure. The regression coefficients would be estimated utilizing the completed interviews only. Then the payment error for each case with missing data would be imputed from the appropriate regression equation. Given these values, the procedure for imputing missing Federal findings might proceed as described in the two preceding approaches.

For each of the procedures considered modifications and extensions would likely be introduced as the work proceeded and results were observed.

We note that imputation procedures (1) and (2) would likely yield reasonably satisfactory results if the cells defined were sufficiently homogeneous that the frequency of the errors in determining eligibility and the magnitude of the issuances were, within cells, similar, except for sampling variability, for the completed cases and the imputed cases.

Table of Contents

4. SUMMARY AND RECOMMENDATIONS

Under current practice the proportion of sample cases not reviewed (including some inactive cases, the number of which cannot be identified from currently available information) averages nearly ten percent over the United States, and has been as high as twenty-five percent in some states. The procedures described and recommended in Section 2 would greatly reduce the frequency of such cases.

We therefore recommend that the only sample cases not subject to review be cases that were not active in the sample month and cases that are dropped from the sample at random to avoid a sample size greater than required.

We recommend that, in calculation of error rates, reviews not completed (whether by the state or Federal reviewer) be handled by the

Table of Contents

and allotment errors in such cases, and whether explicit imputation is worthwhile for these remaining cases, and, if so, how imputation should be done. However, if the nonresponse rate is sufficiently low after implementation of the procedures proposed here, sensitivity analysis may show that such investigation is not necessary. If an imputation procedure is adopted, the Federal sample should also include a subsample of the state cases for which an imputation has been made.

In summary, we anticipate that adoption of the recommended procedures would mean that many or most of the reviews currently recorded as not subject to review or not completed would be completed by reference to the case records and to other sources of information. However, presumably some uncompleted reviews of active cases would remain. Under these circumstances one of the following alternatives for estimating the various quantities in the regression estimator ($\bar{X}$, $\bar{x}$, $\bar{y}$, and b) is recommended, the choice depending on what is observed, as indicated.

(1) Estimate each of the quantities implicitly from the values provided by the completed cases in the state sample, and in the Federal subsample, as is currently done. If there is a small enough number of remaining not completed cases, sensitivity analyses will show this to be an acceptable alternative.

(2) If the sensitivity analyses show that the estimates might still be sensitive to the treatment of the remaining uncompleted cases, imputation for such cases should be done for both the state sample cases and the Federal subsample cases. The choice of imputation procedures will depend on analysis of the remaining uncompleted cases. These analyses cannot be done until results of the new procedures are in effect, and their consequences can be observed.

(3) The Appendix shows that it is not likely to be cost effective for the Federal reviewer to arrive at a determination for each such remaining case in the state sample, by whatever means, and then include each of them in the state sample as completed reviews, and an appropriate subsample of them in the Federal subsample.

(4) Another alternative is for the Federal reviewer to examine only the Federal subsample of such remaining

Table of Contents

cases and arrive at a plausible determination for each case by whatever means. These determinations could then be used in estimating $\bar{y}$, but not in estimating $\bar{X}$, $\bar{x}$, and b. This procedure is acceptable and recommended if the Federal reviewer could arrive at acceptable determinations for each such remaining case, or for some of them. An evaluation of this procedure would be needed before its adoption.

Final evaluation and choices among these alternatives must await the result of the adoption of new procedures and analysis of the remaining not completed cases. We recommend that alternative (1) above be adopted pending such additional evaluation. The "concluding comments" in Section 2 provide explicit details on the estimation procedure under this alternative.

Table of Contents

Table 1. Frequency by state disposition codes, Food Stamps, FY 1986

	Frequency by state disposition code							Relative frequency by state disposition code					
State	Total	1	2	4	5	6	7	1	2	4	5	6	7
AK	331	297	25	7	1	0	1	0.8973	0.0755	0.0211	0.0030	0.0000	0.0030
AL	2293	2239	33	15	5	0	1	0.9765	0.0144	0.0065	0.0022	0.0000	0.0004
AR	1316	1288	19	3	2	1	3	0.9787	0.0144	0.0023	0.0015	0.0008	0.0023
AZ	2740	2361	228	70	42	10	29	0.8617	0.0832	0.0255	0.0153	0.0036	0.0106
CA	2596	2276	204	73	14	1	28	0.8767	0.0786	0.0281	0.0054	0.0004	0.0108
CO	1385	1305	62	9	6	1	2	0.9422	0.0448	0.0065	0.0043	0.0007	0.0014
CT	1098	1025	58	14	0	0	1	0.9335	0.0528	0.0128	0.0000	0.0000	0.0009
DC	662	603	30	28	0	0	1	0.9109	0.0453	0.0423	0.0000	0.0000	0.0015
DE	369	350	11	3	0	0	5	0.9485	0.0298	0.0081	0.0000	0.0000	0.0136
FL	2551	2416	88	21	0	0	26	0.9471	0.0345	0.0082	0.0000	0.0000	0.0102
GA	1276	1202	46	19	2	1	6	0.9420	0.0361	0.0149	0.0016	0.0008	0.0047
GQ	319	306	11	0	0	0	2	0.9592	0.0345	0.0000	0.0000	0.0000	0.0063
HI	831	773	49	6	0	1	2	0.9302	0.0590	0.0072	0.0000	0.0012	0.0024
IA	1348	1250	50	40	5	0	3	0.9273	0.0371	0.0297	0.0037	0.0000	0.0022
ID	996	873	52	38	23	1	9	0.8765	0.0522	0.0382	0.0231	0.0010	0.0090
IL	3081	2923	96	45	5	0	12	0.9487	0.0312	0.0146	0.0016	0.0000	0.0039
IN	1320	1214	81	18	6	0	1	0.9197	0.0614	0.0136	0.0045	0.0000	0.0008
KS	993	940	38	12	2	0	1	0.9466	0.0383	0.0121	0.0020	0.0000	0.0010
KY	1737	1635	50	46	0	0	6	0.9413	0.0288	0.0265	0.0000	0.0000	0.0035
LA	1280	1213	48	15	3	0	1	0.9477	0.0375	0.0117	0.0023	0.0000	0.0008
MA	1302	1194	42	60	6	0	0	0.9171	0.0323	0.0461	0.0046	0.0000	0.0000
MD	1314	1211	69	28	3	0	3	0.9216	0.0525	0.0213	0.0023	0.0000	0.0023
ME	1068	1019	46	3	0	0	0	0.9541	0.0431	0.0028	0.0000	0.0000	0.0000
MI	2513	2311	69	102	7	0	24	0.9196	0.0275	0.0406	0.0028	0.0000	0.0096
MN	1376	1245	72	48	8	0	3	0.9048	0.0523	0.0349	0.0058	0.0000	0.0022
MO	2782	2651	108	14	7	0	2	0.9529	0.0388	0.0050	0.0025	0.0000	0.0007
MS	1412	1238	159	3	11	0	1	0.8768	0.1126	0.0021	0.0078	0.0000	0.0007
MT	1057	869	181	6	0	0	1	0.8221	0.1712	0.0057	0.0000	0.0000	0.0009
NC	1271	1217	21	17	9	0	7	0.9575	0.0165	0.0134	0.0071	0.0000	0.0055
ND	333	328	4	1	0	0	0	0.9850	0.0120	0.0030	0.0000	0.0000	0.0000
NE	1454	1373	72	9	0	0	0	0.9443	0.0495	0.0062	0.0000	0.0000	0.0000
NH	551	528	22	1	0	0	0	0.9583	0.0399	0.0018	0.0000	0.0000	0.0000
NJ	2704	2376	199	102	20	1	6	0.8787	0.0736	0.0377	0.0074	0.0004	0.0022
NM	1504	1425	27	31	15	1	5	0.9475	0.0180	0.0206	0.0100	0.0007	0.0033
NV	657	590	59	1	5	0	2	0.8980	0.0898	0.0015	0.0076	0.0000	0.0030
NY	1350	1232	87	26	0	0	5	0.9126	0.0644	0.0193	0.0000	0.0000	0.0037
OH	1414	1293	72	37	8	0	4	0.9144	0.0509	0.0262	0.0057	0.0000	0.0028
OK	1513	1343	152	18	0	0	0	0.8876	0.1005	0.0119	0.0000	0.0000	0.0000
OR	3636	2937	578	75	36	1	9	0.8078	0.1590	0.0206	0.0099	0.0003	0.0025
PA	1335	1235	61	34	3	1	1	0.9251	0.0457	0.0255	0.0022	0.0007	0.0007
RI	1157	1051	57	44	4	0	1	0.9084	0.0493	0.0380	0.0035	0.0000	0.0009
SC	1368	1225	109	34	0	0	0	0.8955	0.0797	0.0249	0.0000	0.0000	0.0000
SD	644	616	15	11	2	0	0	0.9565	0.0233	0.0171	0.0031	0.0000	0.0000
TN	1370	1324	21	12	8	0	5	0.9664	0.0153	0.0088	0.0058	0.0000	0.0036
TX	1320	1258	55	7	0	0	0	0.9530	0.0417	0.0053	0.0000	0.0000	0.0000
UT	635	592	42	1	0	0	0	0.9323	0.0661	0.0016	0.0000	0.0000	0.0000
VA	1290	1193	48	43	1	0	5	0.9248	0.0372	0.0333	0.0008	0.0000	0.0039
VQ	304	298	1	4	1	0	0	0.9803	0.0033	0.0132	0.0033	0.0000	0.0000
VT	423	406	11	5	0	0	1	0.9598	0.0260	0.0118	0.0000	0.0000	0.0024
WA	2872	2614	209	40	5	0	4	0.9102	0.0728	0.0139	0.0017	0.0000	0.0014
WI	2337	2167	123	33	11	0	3	0.9273	0.0526	0.0141	0.0047	0.0000	0.0013
WV	1294	1222	39	22	7	4	0	0.9444	0.0301	0.0170	0.0054	0.0031	0.0000
WY	345	323	15	4	3	0	0	0.9362	0.0435	0.0116	0.0087	0.0000	0.0000
U.S.	74427	68393	4124	1358	296	24	232	0.9189	0.0554	0.0182	0.0040	0.0003	0.0031

Table 1. Frequency by state disposition codes, Food Stamps, FY 1986

	Standard error under null hypothesis						Studentized ratio					
State	1	2	4	5	6	7	1	2	4	5	6	7
AK	0.0150	0.0126	0.0074	0.0035	0.0010	0.0031	-1.443	1.5999	0.3945	-0.276	-0.327	-0.031
AL	0.0057	0.0048	0.0028	0.0013	0.0004	0.0012	10.092	-8.585	-4.188	-1.367	-0.86	-2.303
AR	0.0075	0.0063	0.0037	0.0017	0.0005	0.0015	7.9474	-6.497	-4.328	-1.416	0.8838	-0.545
AZ	0.0052	0.0044	0.0026	0.0012	0.0003	0.0011	-10.98	6.3611	2.8556	9.4408	9.7002	7.0115
CA	0.0054	0.0045	0.0026	0.0012	0.0004	0.0011	-7.876	5.1607	3.7589	1.1462	0.1781	7.0093
CO	0.0073	0.0061	0.0036	0.0017	0.0005	0.0015	3.1784	-1.732	-3.267	0.21	0.8282	-1.117
CT	0.0082	0.0069	0.0040	0.0019	0.0005	0.0017	1.771	-0.375	-1.361	-2.094	-0.595	-1.312
DC	0.0106	0.0089	0.0052	0.0024	0.0007	0.0022	-0.759	-1.135	4.6234	-1.626	-0.462	-0.742
DE	0.0142	0.0119	0.0070	0.0033	0.0009	0.0029	2.0819	-2.149	-1.452	-1.214	-0.345	3.5952
FL	0.0054	0.0045	0.0026	0.0012	0.0004	0.0011	5.2094	-4.617	-3.779	-3.192	-0.907	6.4103
GA	0.0076	0.0064	0.0037	0.0018	0.0005	0.0016	3.0204	-3.023	-0.896	-1.368	0.9177	1.0157
GQ	0.0153	0.0128	0.0075	0.0035	0.0010	0.0031	2.6384	-1.634	-2.435	-1.129	-0.321	1.01
HI	0.0095	0.0079	0.0046	0.0022	0.0006	0.0019	1.191	0.448	-2.375	-1.822	1.4144	-0.367
IA	0.0074	0.0062	0.0036	0.0017	0.0005	0.0015	1.1262	-2.94	3.1348	-0.156	-0.659	-0.587
ID	0.0086	0.0072	0.0042	0.0020	0.0006	0.0018	-4.905	-0.442	4.694	9.5851	1.198	3.351
IL	0.0049	0.0041	0.0024	0.0011	0.0003	0.0010	6.0583	-5.884	-1.51	-2.076	-0.997	0.7744
IN	0.0075	0.0063	0.0037	0.0017	0.0005	0.0015	0.1025	0.9455	-1.251	0.3281	-0.653	-1.538
KS	0.0087	0.0073	0.0042	0.0020	0.0006	0.0018	3.1979	-2.361	-1.451	-0.983	-0.566	-1.193
KY	0.0065	0.0055	0.0032	0.0015	0.0004	0.0013	3.4128	-4.85	2.5648	-2.634	-0.749	0.252
LA	0.0076	0.0064	0.0037	0.0018	0.0005	0.0016	3.7657	-2.801	-1.745	-0.928	-0.643	-1.499
MA	0.0076	0.0063	0.0037	0.0017	0.0005	0.0015	-0.248	-3.652	7.5048	0.3619	-0.648	-2.018
MD	0.0075	0.0063	0.0037	0.0017	0.0005	0.0015	0.3567	-0.459	0.8296	-0.976	-0.651	-0.542
ME	0.0084	0.0070	0.0041	0.0019	0.0005	0.0017	4.2137	-1.763	-3.769	-2.065	-0.587	-1.827
MI	0.0054	0.0046	0.0027	0.0013	0.0004	0.0011	0.1269	-6.125	8.3685	-0.949	-0.9	5.7852
MN	0.0074	0.0062	0.0036	0.0017	0.0005	0.0015	-1.92	-0.5	4.6112	1.0826	-0.666	-0.623
MO	0.0052	0.0043	0.0025	0.0012	0.0003	0.0011	6.5672	-3.825	-5.207	-1.224	-0.947	-2.269
MS	0.0073	0.0061	0.0036	0.0017	0.0005	0.0015	-5.804	9.3944	-4.526	2.2767	-0.675	-1.624
MT	0.0084	0.0070	0.0041	0.0019	0.0006	0.0017	-11.53	16.46	-3.053	-2.054	-0.584	-1.266
NC	0.0077	0.0064	0.0038	0.0018	0.0005	0.0016	5.04	-6.06	-1.297	1.7582	-0.64	1.5287
ND	0.0150	0.0125	0.0073	0.0034	0.0010	0.0031	4.4164	-3.462	-2.078	-1.153	-0.328	-1.02
NE	0.0072	0.0060	0.0035	0.0017	0.0005	0.0015	3.5435	-0.982	-3.435	-2.41	-0.685	-2.132
NH	0.0116	0.0097	0.0057	0.0027	0.0008	0.0024	3.3824	-1.589	-2.882	-1.483	-0.422	-1.313
NJ	0.0052	0.0044	0.0026	0.0012	0.0003	0.0011	-7.664	4.1333	7.5668	2.8251	0.1372	-0.838
NM	0.0070	0.0059	0.0035	0.0016	0.0005	0.0014	4.056	-6.35	0.6855	3.6948	0.7396	0.1442
NV	0.0106	0.0089	0.0052	0.0025	0.0007	0.0022	-1.963	3.8532	-3.203	1.4797	-0.46	-0.034
NY	0.0074	0.0062	0.0036	0.0017	0.0005	0.0015	-0.853	1.451	0.2781	-2.322	-0.66	0.3866
OH	0.0073	0.0061	0.0036	0.0017	0.0005	0.0015	-0.62	-0.738	2.2254	1.0041	-0.675	-0.194
OK	0.0070	0.0059	0.0034	0.0016	0.0005	0.0014	-4.459	7.6599	-1.845	-2.458	-0.699	-2.175
OR	0.0045	0.0038	0.0022	0.0010	0.0003	0.0009	-24.56	27.294	1.0727	5.6755	-0.159	-0.694
PA	0.0075	0.0063	0.0037	0.0017	0.0005	0.0015	0.8255	-1.552	1.9716	-1.004	0.8681	-1.552
RI	0.0080	0.0067	0.0039	0.0019	0.0005	0.0016	-1.314	-0.914	5.0278	-0.281	-0.611	-1.375
SC	0.0074	0.0062	0.0036	0.0017	0.0005	0.0015	-3.179	3.9234	1.826	-2.337	-0.664	-2.068
SD	0.0108	0.0090	0.0053	0.0025	0.0007	0.0022	3.4953	-3.563	-0.221	-0.351	-0.456	-1.419
TN	0.0074	0.0062	0.0036	0.0017	0.0005	0.0015	6.4408	-6.485	-2.624	1.0952	-0.665	0.3536
TX	0.0075	0.0063	0.0037	0.0017	0.0005	0.0015	4.5394	-2.183	-3.513	-2.296	-0.653	-2.032
UT	0.0108	0.0091	0.0053	0.0025	0.0007	0.0022	1.2331	1.1821	-3.139	-1.592	-0.453	-1.409
VA	0.0076	0.0064	0.0037	0.0018	0.0005	0.0016	0.7736	-2.857	4.0487	-1.827	-0.645	0.4889
VQ	0.0157	0.0131	0.0077	0.0036	0.0010	0.0032	3.9181	-3.972	-0.663	-0.19	-0.313	-0.975
VT	0.0133	0.0111	0.0065	0.0031	0.0009	0.0027	3.0806	-2.644	-0.987	-1.3	-0.369	-0.278
WA	0.0051	0.0043	0.0025	0.0012	0.0003	0.0010	-1.72	4.0669	-1.729	-1.904	-0.963	-1.658
WI	0.0056	0.0047	0.0028	0.0013	0.0004	0.0012	1.4753	-0.587	-1.49	0.5606	-0.868	-1.59
WV	0.0076	0.0064	0.0037	0.0017	0.0005	0.0015	3.3516	-3.973	-0.334	0.8188	5.5472	-2.012
WY	0.0147	0.0123	0.0072	0.0034	0.0010	0.0030	1.1776	-0.969	-0.923	1.3925	-0.334	-1.039
U.S.												

Table of Contents

Table 2. Frequancy by state disposition codes, Food Stamps, FY 19

State		Frequency by state disposition code						Relative frequency by state disposition code					
	Total	1	2	4	5	6	7	1	2	4	5	6	7
AK	354	332	16	3	3	0	0	.9379	.0452	.0085	.0085	.0000	.0000
AL	1931	1878	32	17	4	0	0	.9726	.0166	.0088	.0021	.0000	.0000
AR	1328	1287	27	8	4	0	2	.9691	.0203	.0060	.0030	.0000	.0015
AZ	2693	2343	244	68	7	1	30	.8700	.0906	.0253	.0026	.0004	.0111
CA	2608	2293	220	72	8	0	15	.8792	.0844	.0276	.0031	.0000	.0058
CO	1472	1329	113	19	5	0	6	.9029	.0768	.0129	.0034	.0000	.0041
CT	962	899	56	5	0	0	2	.9345	.0582	.0052	.0000	.0000	.0021
DC	642	597	37	6	1	0	1	.9299	.0576	.0093	.0016	.0000	.0016
DE	355	336	14	3	1	0	1	.9465	.0394	.0085	.0028	.0000	.0028
FL	2548	2422	68	36	0	0	22	.9505	.0267	.0141	.0000	.0000	.0086
GA	1284	1210	55	11	1	0	7	.9424	.0428	.0086	.0008	.0000	.0055
GU	331	320	11	0	0	0	0	.9668	.0332	.0000	.0000	.0000	.0000
HI	807	758	42	6	1	0	0	.9393	.0520	.0074	.0012	.0000	.0000
IA	1317	1225	33	40	8	3	8	.9301	.0251	.0304	.0061	.0023	.0061
ID	999	755	169	37	19	7	12	.7558	.1692	.0370	.0190	.0070	.0120
IL	2873	2690	106	52	5	0	20	.9363	.0369	.0181	.0017	.0000	.0070
IN	1300	1207	77	13	0	0	3	.9285	.0592	.0100	.0000	.0000	.0023
KS	1011	973	34	4	0	0	0	.9624	.0336	.0040	.0000	.0000	.0000
KY	1750	1639	50	57	0	0	4	.9366	.0286	.0326	.0000	.0000	.0023
LA	1266	1214	30	12	8	1	1	.9589	.0237	.0095	.0063	.0008	.0008
MA	1306	1205	53	41	4	0	3	.9227	.0406	.0314	.0031	.0000	.0023
MD	1373	1244	59	51	4	0	15	.9060	.0430	.0371	.0029	.0000	.0109
ME	967	931	33	3	0	0	0	.9628	.0341	.0031	.0000	.0000	.0000
MI	2475	2286	74	68	2	0	45	.9236	.0299	.0275	.0008	.0000	.0182
MN	1372	1261	58	48	3	0	2	.9191	.0423	.0350	.0022	.0000	.0015
MO	2540	2419	96	15	7	1	2	.9524	.0378	.0059	.0028	.0004	.0008
MS	1416	1241	162	3	10	0	0	.8764	.1144	.0021	.0071	.0000	.0000
MT	1038	854	139	43	2	0	0	.8227	.1339	.0414	.0019	.0000	.0000
NC	1222	1167	18	25	4	0	8	.9550	.0147	.0205	.0033	.0000	.0065
ND	378	371	5	1	1	0	0	.9815	.0132	.0026	.0026	.0000	.0000
NE	809	761	39	9	0	0	0	.9407	.0482	.0111	.0000	.0000	.0000
NH	441	425	16	0	0	0	0	.9637	.0363	.0000	.0000	.0000	.0000
NJ	2545	2238	170	99	15	3	20	.8794	.0668	.0389	.0059	.0012	.0079
NM	1255	1202	21	20	3	6	3	.9578	.0167	.0159	.0024	.0048	.0024
NV	638	539	90	1	7	0	1	.8448	.1411	.0016	.0110	.0000	.0016
NY	1297	1171	70	43	0	0	13	.9029	.0540	.0332	.0000	.0000	.0100
OH	1302	1194	57	35	12	0	4	.9171	.0438	.0269	.0092	.0000	.0031
OK	1469	1407	48	13	0	1	0	.9578	.0327	.0088	.0000	.0007	.0000
OR	2681	2435	143	67	25	4	7	.9082	.0533	.0250	.0093	.0015	.0026
PA	1280	1190	57	30	3	0	0	.9297	.0445	.0234	.0023	.0000	.0000
RI	1086	1008	50	27	1	0	0	.9282	.0460	.0249	.0009	.0000	.0000
SC	1354	1226	95	30	2	0	1	.9055	.0702	.0222	.0015	.0000	.0007
SD	646	615	24	5	2	0	0	.9520	.0372	.0077	.0031	.0000	.0000
TN	1329	1279	18	15	9	1	7	.9624	.0135	.0113	.0068	.0008	.0053
TX	1320	1241	57	22	0	0	0	.9402	.0432	.0167	.0000	.0000	.0000
UT	675	620	50	5	0	0	0	.9185	.0741	.0074	.0000	.0000	.0000
VA	1286	1213	49	15	2	0	7	.9432	.0381	.0117	.0016	.0000	.0054
VI	308	303	0	1	2	0	2	.9838	.0000	.0032	.0065	.0000	.0065
VT	414	402	10	2	0	0	0	.9710	.0242	.0048	.0000	.0000	.0000
WA	2598	2384	165	32	7	0	10	.9176	.0635	.0123	.0027	.0000	.0038
WI	2285	2115	121	36	7	0	6	.9256	.0530	.0158	.0031	.0000	.0026
WV	1271	1213	27	26	4	1	0	.9544	.0212	.0205	.0031	.0008	.0000
WY	345	311	16	9	8	1	0	.9014	.0464	.0261	.0232	.0029	.0000
Total	70552	65178	3524	1309	221	30	290	.9238	.0499	.0186	.0031	.0004	.0041

Table 2. Frequency by state disposition codes, Food Stamps, FY 19

State	Standard error under null hypothesis						Studenti					
	1	2	4	5	6	7	1	2	4	5	6	7
AK	.0141	.0116	.0072	.0030	.0011	.0138	0.99	-0.41	-1.41	1.80	-0.39	-0.30
AL	.0060	.0050	.0031	.0013	.0005	.0059	8.07	-6.73	-3.18	-0.83	-0.91	-0.70
AR	.0073	.0060	.0037	.0015	.0006	.0071	6.22	-4.95	-3.38	-0.08	-0.75	-0.37
AZ	.0051	.0042	.0026	.0011	.0004	.0050	-10.52	9.69	2.58	-0.50	-0.14	1.41
CA	.0052	.0043	.0026	.0011	.0004	.0051	-8.59	8.07	3.43	-0.06	-1.05	0.32
CO	.0069	.0057	.0035	.0015	.0005	.0067	-3.03	4.72	-1.61	0.18	-0.79	-0.01
CT	.0086	.0070	.0044	.0018	.0007	.0083	1.25	1.18	-3.07	-1.74	-0.64	-0.24
DC	.0105	.0086	.0053	.0022	.0008	.0102	0.58	0.89	-1.73	-0.71	-0.52	-0.25
DE	.0141	.0116	.0072	.0030	.0011	.0137	1.61	-0.91	-1.41	-0.11	-0.39	-0.09
FL	.0053	.0043	.0027	.0011	.0004	.0051	5.08	-5.39	-1.66	-2.83	-1.04	0.88
GA	.0074	.0061	.0038	.0016	.0006	.0072	2.50	-1.17	-2.65	-1.51	-0.74	0.19
GQ	.0146	.0120	.0074	.0031	.0011	.0142	2.94	-1.40	-2.50	-1.02	-0.38	-0.29
HI	.0093	.0077	.0048	.0020	.0007	.0091	1.65	0.27	-2.34	-0.96	-0.59	-0.45
IA	.0073	.0060	.0037	.0015	.0006	.0071	0.86	-4.15	3.18	1.91	3.26	0.28
ID	.0084	.0069	.0043	.0018	.0007	.0082	-20.03	17.30	4.33	8.99	10.09	0.97
IL	.0049	.0041	.0025	.0010	.0004	.0012	2.52	-3.21	-0.18	-1.34	-1.11	2.39
IN	.0074	.0060	.0037	.0015	.0006	.0072	0.63	1.54	-2.29	-2.02	-0.74	-0.25
KS	.0083	.0069	.0042	.0018	.0006	.0081	4.62	-2.38	-3.44	-1.78	-0.66	-0.51
KY	.0063	.0052	.0032	.0013	.0005	.0062	2.01	-4.11	4.35	-2.34	-0.86	-0.29
LA	.0075	.0061	.0038	.0016	.0006	.0073	4.71	-4.29	-2.39	2.03	0.63	-0.46
MA	.0073	.0060	.0037	.0015	.0006	.0072	-0.16	-1.55	3.44	-0.05	-0.75	-0.25
MD	.0072	.0059	.0036	.0015	.0006	.0070	-2.48	-1.19	5.10	-0.15	-0.76	0.98
ME	.0085	.0070	.0043	.0018	.0007	.0083	4.57	-2.26	-3.56	-1.74	-0.64	-0.49
MI	.0053	.0044	.0027	.0011	.0004	.0052	-0.04	-4.58	3.29	-2.07	-1.03	2.70
MN	.0072	.0059	.0036	.0015	.0006	.0070	-0.66	-1.31	4.51	-0.63	-0.76	-0.38
MO	.0053	.0043	.0027	.0011	.0004	.0051	5.42	-2.81	-4.72	-0.34	-0.08	-0.65
MS	.0070	.0058	.0036	.0015	.0005	.0069	-6.73	11.13	-4.58	2.65	-0.78	-0.60
MT	.0082	.0068	.0042	.0017	.0006	.0080	-12.28	12.42	5.46	-0.70	-0.66	-0.51
NC	.0076	.0062	.0039	.0016	.0006	.0074	4.11	-5.65	0.49	0.09	-0.72	0.33
ND	.0136	.0112	.0069	.0029	.0011	.0133	4.23	-3.28	-2.29	-0.17	-0.40	-0.31
NE	.0093	.0077	.0047	.0020	.0007	.0091	1.81	-0.23	-1.57	-1.59	-0.59	-0.45
NH	.0126	.0104	.0064	.0027	.0010	.0123	3.16	-1.32	-2.89	-1.18	-0.43	-0.33
NJ	.0053	.0043	.0027	.0011	.0004	.0051	-8.45	3.90	7.61	2.49	1.84	0.73
NM	.0075	.0061	.0038	.0016	.0006	.0073	4.53	-5.40	-0.69	-0.47	7.48	-0.24
NV	.0105	.0086	.0053	.0022	.0008	.0102	-7.52	10.57	-3.18	3.54	-0.52	-0.25
NY	.0074	.0060	.0037	.0016	.0006	.0072	-2.85	0.66	3.90	-2.02	-0.74	0.82
OH	.0074	.0060	.0037	.0015	.0006	.0072	-0.92	-1.02	2.23	3.93	-0.74	-0.14
OK	.0069	.0057	.0035	.0015	.0005	.0068	4.91	-3.04	-2.76	-2.15	0.48	-0.61
OR	.0051	.0042	.0026	.0011	.0004	.0050	-3.04	0.81	2.47	5.74	2.68	-0.30
PA	.0074	.0061	.0038	.0016	.0006	.0072	0.79	-0.89	1.29	-0.50	-0.74	-0.57
RI	.0080	.0066	.0041	.0017	.0006	.0079	0.54	-0.59	1.54	-1.30	-0.68	-0.52
SC	.0072	.0059	.0037	.0015	.0006	.0070	-2.55	3.41	0.98	-1.09	-0.76	-0.48
SD	.0104	.0086	.0053	.0022	.0008	.0102	2.70	-1.49	-2.04	-0.02	-0.52	-0.40
TN	.0073	.0060	.0037	.0015	.0006	.0071	5.30	-6.09	-1.96	2.37	0.58	0.16
TX	.0073	.0060	.0037	.0015	.0006	.0071	2.24	-1.13	-0.51	-2.04	-0.75	-0.58
UT	.0102	.0084	.0052	.0022	.0008	.0100	-0.52	2.88	-2.15	-1.46	-0.54	-0.41
VA	.0074	.0061	.0038	.0016	.0006	.0072	2.62	-1.95	-1.83	-1.01	-0.74	0.18
VQ	.0151	.0124	.0077	.0032	.0012	.0147	3.97	-4.02	-1.99	1.06	-0.36	0.16
VT	.0130	.0107	.0066	.0027	.0010	.0127	3.62	-2.41	-2.07	-1.14	-0.42	-0.32
WA	.0052	.0043	.0026	.0011	.0004	.0051	-1.19	3.17	-2.36	-0.40	-1.05	-0.05
WI	.0055	.0046	.0028	.0012	.0004	.0054	0.32	0.66	-0.99	-0.06	-0.99	-0.27
WV	.0074	.0061	.0038	.0016	.0006	.0073	4.10	-4.70	0.50	0.01	0.63	-0.57
WY	.0143	.0117	.0073	.0030	.0011	.0139	-1.57	-0.30	1.04	6.67	2.23	-0.30
Total												

Table of Contents

Appendix

Sensitivity of the correlation and regression coefficients to added cases in the state sample and the Federal subsample

It has been suggested that when cases not completed by the state review are completed by the Federal review as a result of the Federal review of all such cases in the state sample, the Federal error finding should be used as if it were a state finding in the computation of $\bar{X}$, $\bar{x}$, and b and that the appropriate subsample of such cases be used in the computation of $\bar{y}$. This procedure will increase the correlation between $\bar{y}$ and $\bar{x}$, and will also increase n' and n, and will therefore reduce the variance of $\bar{y}'$, the regression estimate of the average error per case. The question arises whether the cost of the Federal review of all cases in the state sample is justified by the resulting reduction in the variance of the error rate.

The variance of $\bar{y}$ is given approximately by

$$\frac{\sigma_Y^2}{n'}\left[1 - \rho^2(1-f)\right],$$

where n' is the size of the Federal subsample, f is the subsampling fraction, σ_Y^2 is the variance of Federal error findings, and ρ is the correlation coefficient. The effect of the added cases is to decrease the value of the expression in brackets and to increase the value of n'. Each case added to the Federal subsample requires the Federal review of approximately an additional 1/f of the state sample cases that must have a Federal review; this is (1–f)/f more cases than would be required if a Federal review is made only for cases in the initial subsample of completed state cases. For example, if f=1/6 then 5 cases must be reviewed (on the average) for inclusion in $\bar{X}$ to add a case in the subsample. The additional cost of the Federal review is then c(1–f)/f where c denotes the unit cost of a Federal review. The relative increase in the cost of Federal review is therefore

$$(1-f)/fn'.$$

Table of Contents

To get an impression of the sensitivity of the correlation r and the regression coefficient b when a case is added, we used the Federal subsample for three states in FY 1987. One was a a state with a large sample, one was a state with a small sample and a high correlation, and one was a state with a small sample and a lower correlation. For each state, we calculated the correlation after adding a case to the subsample at a number of distances from the mean error per case. The results are summarized below.

Experiment No. 1:

$n' = 386$ $f = .3319$ $\rho = .9586$ $b = 1.0066$

$\bar{x} = 12.65$ $s_x = 34.33$ $\bar{y} = 14.30$ $s_y = 36.05$

Added point

x	y	ρ	b	Relative variance reduction
13	14	.9586	1.0066	.0000
50	52	.9587	1.0066	.0003
80	82	.9590	1.0066	.0013
110	112	.9594	1.0066	.0027
140	143	.9600	1.0068	.0046
170	173	.9606	1.0067	.0066

Relative cost increase = .0052

Table of Contents

Experiment No. 2:

$n' = 226$ $f = .3767$ $\rho = .9407$ $b = .9934$

$\bar{x} = 3.7257$ $s_x = 29.79$ $\bar{y} = 4.6239$ $s_y = 31.46$

Added point

x	y	ρ	b	Relative variance reduction
4	5	.9407	.9934	.0000
30	31	.9409	.9935	.0005
60	61	.9416	.9936	.0024
90	90	.9427	.9934	.0052
120	120	.9442	.9935	.0092
150	150	.9420	.9893	.0140

Relative cost increase = .0073

Experiment No. 3:

$n' = 321$ $f = .3847$ $\rho = .8982$ $b = .9216$

$\bar{x} = 5.0969$ $s_x = 31.83$ $\bar{y} = 5.1189$ $s_y = 32.66$

Added point

x	y	ρ	b	Relative variance reduction
5	5	.8982	.9216	.0000
35	33	.8986	.9218	.0009
65	60	.8995	.9219	.0029
95	88	.9012	.9225	.0066
125	116	.9034	.9235	.0115
160	148	.9076	.9251	.0208

Relative cost increase = .0070

Table of Contents

In these examples it is seen that the relative increase in the Federal review cost is greater in magnitude than the relative decrease in variance unless the added case is distant from the mean by more than about three times the standard deviation. Although the distribution of errors is skewed to the right, such large deviations occur with low probabilities, so that the probability that the added cost will be less than the added gain is small. It is also noted that even for the states with small samples both the cost and the gain are small.

Theory

Let ρ be the population correlation of $\bar{x}$ and $\bar{y}$. Let ξ and η be the sample means when a random value from the distribution of y is appended to the sample means of size n. That is,

$$\xi = \frac{n\bar{x} + y}{n+1} = \frac{n}{n+1}\bar{x} + \frac{1}{n+1}y$$

$$\eta = \frac{n\bar{y} + y}{n+1} = \frac{n}{n+1}\bar{y} + \frac{1}{n+1}y$$

Hence

$$\mathrm{Var}(\xi) = \frac{n^2}{(n+1)^2}\frac{\sigma_X^2}{n} + \frac{1}{(n+1)^2}\sigma_Y^2 = \frac{n}{(n+1)^2}\sigma_X^2 + \frac{1}{(n+1)^2}\sigma_Y^2$$

Table of Contents

Let us define

$$c = \frac{\sigma_X}{\sigma_Y}.$$

Note that in the Food Stamp context $c < 1$. It follows that the correlation between ξ and η is

$$\rho_{\xi\eta} = \frac{\frac{1}{(n+1)^2}\left[nc\rho + 1\right]}{\frac{1}{(n+1)^2}\sqrt{\left[nc^2+1\right]\left[n+1\right]}} \; \frac{\sigma_Y^2}{\sigma_Y^2}$$

$$= \frac{nc\rho + 1}{\sqrt{\left[nc^2+1\right]\left[n+1\right]}}$$

Thus, for example, if $c=1$, $n=400$ and $\rho=0.9$, the correlation is increased to .9002.

The discussion above has assumed that the added case is a random selection from the y-distribution. I now consider that the added case provides a variable z from a distribution whose mean is $Ez = k_1 Ey$ and whose variance is $\sigma_Z^2 = k_2\sigma_Y^2$. Then

$$\mathrm{Var}(\xi) = \frac{n^2}{(n+1)^2}\frac{\sigma_X^2}{n} + \frac{1}{(n+1)^2}\sigma_Z^2 = \frac{n}{(n+1)^2}\sigma_X^2 + \frac{k_2}{(n+1)^2}\sigma_Y^2$$

$$\mathrm{Var}(\eta) = \frac{n^2}{(n+1)^2}\frac{\sigma_Y^2}{n} + \frac{1}{(n+1)^2}\sigma_Z^2 = \frac{n}{(n+1)^2}\sigma_Y^2 + \frac{k_2}{(n+1)^2}\sigma_Y^2$$

Also

$$\mathrm{Cov}(\xi,\eta) = \frac{n^2}{(n+1)^2}\mathrm{Cov}(\bar{x},\bar{y}) + \frac{1}{(n+1)^2}\sigma_Z^2$$

$$= \frac{n}{(n+1)^2}\sigma_{XY} + \frac{1}{(n+1)^2}\sigma_Z^2$$

$$= \frac{n}{(n+1)^2}\rho\,\sigma_X\sigma_Y + \frac{k_2}{(n+1)^2}\sigma_Y^2$$

Table of Contents

Hence the correlation is now

$$\rho_{\xi\eta} = \frac{nc\rho + k_2}{\sqrt{\left[nc^2 + k_2\right]\left[n + k_2\right]}} .$$